RNA-Seq workflow

Phase 1: developing a workflow and preprocessing raw reads

Andrew Ndhlovu

Project SeaStore
Department of Botany and Zoology
Stellenbosch University
https://github.com/PiscatorX/RNA-Seq-devs

June 26, 2023

Overview of plan of workflow development

Phase I

Check original (raw) data quality

Pre-process reads

- Trim adapter remnants
- Trim low quality bases (Phred score ≤ 25)
- Remove reads ≤ 20nt

Recheck data quality

Phase II

Generate gene/transcript level counts

- Align reads to reference genome, or
- Generate estimated counts using pseudo-alignment approaches

Tabulate overall summary statistics (depends on method used above)

Phase III

Perform quality checks on count data

- Check for outliers among replicates from same set
- Filter out any genes with counts below selected threshold

Perform statistical analysis to find differentially expressed genes

https://h3abionet.github.io/H3ABionet-SOPs/RNA-Seq-1-2.html